A probabilistic risk approach for the collision detection of multi-ships under spatiotemporal movement uncertainty

Xuri Xin[a,1], Kezhong Liu[a,b,d,1], Zaili Yang[c*], Jinfen Zhang[d,e], Xiaolie Wu[a]

[a] *School of Navigation, Wuhan University of Technology, Wuhan, Hubei, China*
[b] *Hubei Key Laboratory of Inland Shipping Technology, Wuhan, Hubei, China*
[c] *Liverpool Logistics, Offshore and Marine (LOOM) Research Institute, Liverpool John Moores University, Liverpool, UK*
[d] *National Engineering Research Center for Water Transport Safety, Wuhan University of Technology, Wuhan, Hubei, China*
[e] *Intelligent Transportation System Center (ITSC), Wuhan University of Technology, Wuhan, Hubei, China*

* Corresponding author.
E-mail address: Z.Yang@ljmu.ac.uk (Z. Yang).

[1] The first two authors contribute equally to this work as the co-first authors.

Abstract: It is vital to analyse ship collision risk for preventing collisions and improving safety at sea. The state-of-the-art of ship collision risk analysis focuses on encountering conflict between ship pairs, subject to a strong assumption of the ships having no/little spatiotemporal motion uncertainty. This paper proposes a probabilistic conflict detection approach to estimate potential collision risk of various multi-vessel encounters, in which the spatiotemporal-dependent patterns of ship motions are newly taken into account through quantifying the trajectory uncertainty distributions using AIS data. The estimation accuracy and efficiency are assured by employing a two-stage Monte Carlo simulation algorithm, which provides the quantitative bounds on the approximation accuracy and allows for a fast estimation of conflict criticality. Several real experiments are conducted using the AIS-based trajectory data in Ningbo-Zhoushan Port to demonstrate the feasibility and superiority of the proposed new approach. The results show that it enables the effective detection of collision risk timely and reliably in a complicated dynamic situation. They therefore provide valuable insights on ship collision risk prediction as well as the formulation of risk mitigation measures.

Keywords: Maritime safety, Ship collision risk, Probabilistic conflict detection, Multi-ship spatiotemporal movements, Uncertainty time-dependent analysis, AIS data.

1. Introduction

Maritime transportation plays a significant role in global economic development. However, the growing shipping traffic volume over the past decades results in high maritime traffic densities or complexity [1], particularly in the waters near ports. It makes ship collisions appearing among the most frequently occurring maritime accident types. In particular, due to their high traffic volumes, such water areas as the Singapore Strait, the Ningbo-Zhoushan Port, and the Northern Baltic Sea are exposed to an extremely complicated traffic situation, in which heavy traffic, high maritime transportation dynamics and changeable vessel motion behaviors often occur. This leads to increasing concerns on the incompetency of traditional risk approaches to maritime traffic risks and challenging demands on new capable models for ship collision risk perception.

In response to such concerns, a variety of methods have been developed for the quantitative analysis of ship collision risk (e.g. [2]–[8]), including theoretical collision risk modelling, probability and consequence assessment, and collision risk estimation, etc. They provide a quantitative basis for implementing ship collision risk mitigation strategies. In recent years, one class of collision risk estimation methods that detect the potential dangerous encounter events from Automatic Identification System (AIS) data using the concepts like “conflict” [9]–[11] or “near miss” [12]–[14] has attracted much research interest. However, the majority of them estimate the collision risk with a strong assumption that the engaged ships will keep the observed velocity in the near future or the ship trajectories can be accurately predicted in advance, overlooking the effects of the dynamic and uncertain characteristics of ship motions. This strong assumption often results in inaccurate conflict or near-miss assessment in reality, especially under highly complicated and dynamic traffic situations. This is because some ships may take one or more turning maneuvers during the encountering process constrained by their navigation plans or water geometry, etc. Moreover, it is extremely hard to accurately predict the ship trajectories due to the uncertainty by various influencing variants such as environmental, physical and human factors. Furthermore, the prediction uncertainty and/or errors are prone to increase gradually over time [15], [16]. A ship-pair encounter scenario is illustrated in Fig. 1, in which the spread of the probability ellipses representing the predicted position uncertainty grows in time. Thus, the performance of these models on risk analysis and prediction becomes questionable and arguable in some practical cases. Another research gap that needs to be addressed in urgency is that most of the current ship collision risk research targets on ship-pairs, and fails to model multi-ship encounter scenarios. It therefore hinders their applications in congested waters where multi-ship encounters are frequently occurring [17]. Consequently, a dynamic risk estimation model that can incorporate the spatial-temporal motion uncertainty of multi-ships becomes essential in order to realize the real-time and accurate evaluation of ship collision risks under high uncertainty.

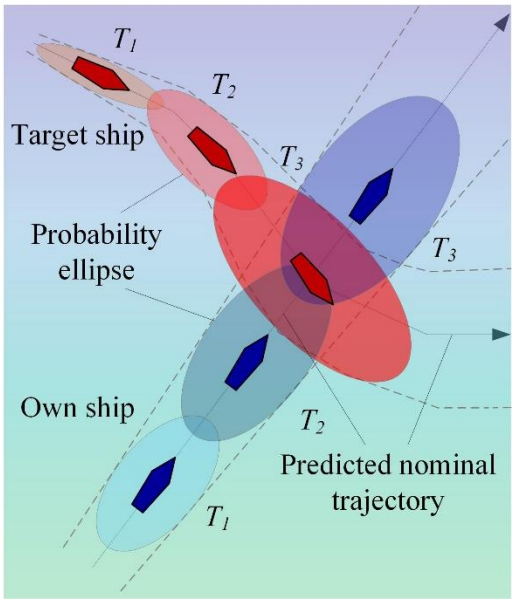


Fig. 1. Illustration of a ship-pair encounter.

This paper aims to develop a probabilistic risk approach for ship conflict detection (CD) that 1) is adaptive to multi-ship encounters in high traffic waters and 2) can take into account the effects of uncertainty (or potential variations) inherent to the spatial-temporal motion of vessels. By doing so, we propose a novel CD approach from a probabilistic risk viewpoint, where an AIS data-driven analysis is conducted to extract the engaged ship trajectory uncertainty distributions. One supporting metric that characterizes the conflicts among ship encounters is first introduced based on a classical ship domain model to measure the criticality of a conflict. Then a ship motion model that incorporates the information on the ship navigation plans and the trajectory uncertainty distributions is newly constructed to predict the ship trajectories in the look-ahead time horizon. Compared to the empirical modelling of the probability density functions (PDFs) of trajectory uncertainty components as Gaussian distributions [15], [16], [18]–[20], this work pioneers to identify the time-dependent position and course uncertainty patterns by mining the trajectory information from historical AIS data. Based on the predicted ship trajectories, the conflict criticality for multiple ships is estimated. Given that the addition of ship motion uncertainty in the model makes the computation of conflict probability very costly, a two-stage Monte Carlo (MC) simulation algorithm is designed and incorporated to efficiently estimate the conflict criticality, while providing the quantitative bounds to ensure the estimation accuracy. The performance of the proposed approach is finally experimentally validated by using real AIS-based trajectory data in port.

More specifically, the rest of the paper is organized as follows. Section 2 provides an overview of the existing collision risk research in the fields of marine and air traffic studies. In Section 3, the proposed probabilistic CD approach is introduced in detail, including the conflict criticality measure, ship motion modelling and conflict

probability estimation. In section 4, the spatiotemporal uncertainty patterns of ship trajectories are extracted and fitted, and the effectiveness and applicability of the proposed models are tested and demonstrated using real data based experiments. Conclusions and directions for future studies are summarized in Section 5.

2. Literature Review

2.1 Ship collision risk analysis

Ship collision risk has long been an active research area in the field of marine traffic management. A growing number of researchers have been working on quantifying ship collision probabilities and/or risk, taking different approaches and from different perspectives. They are at large categorized into the groups of statistical analysis of historical maritime accidents, risk modelling and analysis, and collision risk detection and estimation.

Statistical analysis of historical maritime accidents is one of the fundamental methods to identify the relations between collision frequencies (and/or damage consequences) and risk factors (e.g. vessel attributes, environmental factors, human behaviors, technical failures and traffic situations) [e.g. 21]. Information such as accident databases and accident investigation reports is used to support these studies [2]. To derive knowledge about which factors are highly associated with the ship collision risks, some techniques such as correspondence analysis, logistic regression and stochastic process analysis are employed to predict the probabilities of event-related and environment-dependent accidents [22]–[24], providing insights on the conditions under which maritime traffic accidents may occur. Unfortunately, this type of research strongly relies on historical accident data [25], which sometimes may not be fully available. When concentrating on small research sea/water areas, the occurrence frequency of the collision accidents is usually rare and is often not sufficient for supporting sensible statistical analysis [3]. This issue becomes severe when focusing on investigating and revealing the influence of a large number of the known risk factors influencing ship collision accidents. Therefore, it is essential to adopt additional sources of information to understand the forming mechanism of ship collision risks.

Compared with accident statistical analysis, collision risk modelling integrates multiple sources of information including expert knowledge, historical data, and computer simulation results, etc. It involves two important components in terms of the frequency or probability of ship collision accidents and associated potential consequences. The methods used in this category consist of Macduff [26], Pedersen [27], Fault tree [28], [29], Bayesian networks [30]–[32], and ordered probit model [33], etc. They contribute to the identification of the contributing factors, estimation of the accident causation probability and analysis of their inter-relationships. The findings aid maritime surveillance operators to understand the possible collision frequencies and consequences. Nevertheless, this group of

research provides little value for real-time collision risk detection and collision warning.

Recently, the rapid development of AIS improves the affordability of ship trajectory data acquisition, storage and processing infrastructure and provides great data sources (e.g. AIS data) for vessel collision risk detection and analysis [7]. Various measurable definitions and indices, such as traffic conflict [9], [34], collision candidate [35], [36], near miss [12], [13], and near collision [37] are applied to detect and characterize collision risks. One of the widely used collision risk detection methods is the ship domain-based methods. A ship domain is an important feature of maritime traffic that suggests a region around the ship that needs to be kept clear from other ships or obstacles. A large number of ship domains have been developed for collision risk estimation and analysis, taking into account different shapes (e.g., circular and elliptical [38], [39], quaternion [40], [41], fuzzy [41], [42], projected [43], and risk-based [44]–[46] domains), factors (e.g., ship attributes, environmental conditions, traffic situations, ship maneuverability and knowledge of navigators [47]–[50]), and methodologies (e.g., empirical, knowledge-based and analytical domains [50]–[53]). These models work well for quantitatively examining collision candidates and identifying collision risk hot spots in terms of the intrusion or overlaps of ship domains [10], [11], [54]. However, they are not suitable for performing real-time CD directly. Integrating these models with some trajectory prediction techniques presents a promising research direction for an appropriate solution to tackle this issue.

Another representative category of collision risk analysis is synthetic index methods. This type of methods constructs mathematical or black-box models using various techniques (e.g., Fuzzy Logic [55], Analytical Hierarchical Process [56], Support Vector Machine [57], Multilayer Perceptron [58], and Dempster-Shafer evidence theory [59]) to integrate the factors which can reveal spatial-temporal proximity of approaching ships. Two commonly used indices, namely Distance to Closest Point of Approach (DCPA) and Time to CPA (TCPA), were initially adopted to measure the risk. After that, more influencing factors such as distance, bearing, relative speed, ship maneuverability, stability conditions, and ship domain [56]–[58], [60]–[63] were further incorporated into the models for risk evaluation. To adapt different encounter scenarios and environmental conditions, some models [64] utilized up to 16 indices to estimate the collision risk. However, the models that considered the CPA-based indices imply that they assume the encountering ships will sail linearly without changes of their heading and speeds, and consequently, result in inaccurate collision risk estimation, especially for complicated dynamic traffic situations. Despite there are some research [12], [13], [65], [66] utilizing other indices except for DCPA and TCPA, none of them can cope with the dynamic characteristics of vessel motion, e.g. ship turning maneuvers.

In addition, one type of collision risk detection methods related to dangerous region becomes emerging in the field of maritime traffic. This type of methods performs well in collecting the set of own ship’s speed and/or course that

will cause collision risks with target ships, including Collision Threat Parameter Area (CTPA) [67], Projected Obstacle Area (POA) [68], and Velocity Obstacle (VO) [69], etc. In these studies, VO has become an overwhelming choice for collision estimation due to its simple implementation and robust performance in finding proper collision-free solutions. According to the research in [69]–[71], the VO algorithm accepts the non-linear ship motion assumptions and can be combined with ship maneuverability to detect collision dangers. The problem is however that this algorithm requires the integration of some simple risk measure models (e.g. circular ship domain) to reduce the computational costs of mapping the spatiotemporal proximity between ships to the velocity space. As a result, it has exposed a weakness in terms of reflecting the difference of collision risk in different ship encounter scenarios.

The above literature on ship collision risk reveals valuable insights for ship owners and maritime authorities to propose and implement effective risk mitigation measures. However, most of these collision risk estimation methods focus on the collision risk of ship-pairs and ignore the effects of the ship dynamic and uncertain movements. Only very few studies [17], [66], [72] attach importance to the collision risk of multiple ships, but they work on relatively ideal conditions, such as having regular or constant moving speeds, no current and weather influence on ship motions. In practice, the dynamics and uncertainty of ship potential movements are inevitable and subject to the influence of navigation plan or intention, environmental disturbances, and technical errors, etc. The rare multi-vessel risk estimation studies, together with the ship motion uncertainty, suggest that developing a new collision risk estimation model incorporating spatiotemporal movement uncertainty of multiple vessels is necessary and beneficial.

2.2 Probabilistic conflict detection considering traffic motion uncertainty

In general, the currently available CD methods can be mainly divided into two classes: deterministic and probabilistic. The deterministic approach projects the current state into the future along a single trajectory, i.e., the future trajectories are assumed to be fully known in advance. This class of methods is simple and straightforward but overlooks the trajectory deviations caused by various sources of uncertainty. Currently, the majority of the collision risk estimation research in the maritime sector falls into this category. The probabilistic approach describes the potential variations of the predicted trajectories by using PDFs, from which the probability of a conflict is computed. Thus, this approach can better reflect the reality and is more effective to estimate collision risks under a high level of uncertainty. As the probabilistic CD is a relatively emerging topic under development in the maritime transportation domain, the relevant research in all traffic and transport fields (e.g. air) is critically analyzed.

Probabilistic CD modelling faces several challenges when being applied in a real environment. The first challenge is relating to the computation of the conflict probability based on the PDFs of vessel future trajectories. Conflict probability computation methods in the literature are categorized into Analytical approximations [73]–[76], Monte

Carlo (MC) simulation [20], [77], [78], Gridding methods [25], [70], [79], and Markov chain approximations [80]. Each type of methods has its own strengths and weaknesses (see Table 1). In these methods, the MC simulation is the least restrictive [81]. It enables the non-stationary and non-Gaussian processes, any dependencies and scenarios involving multiple ships to be modelled and evaluated. In addition, it also allows the combination with any types of conflict measure models [82]. Hence, it is used in this paper to perform the conflict probability computation task for multi-ship scenarios with a complex conflict measure model. However, it is recognized that the assessment through MC simulation tends to be computationally intensive due to the slow converges in the process of obtaining accurate results [83]. It is therefore of great importance to improve the efficiency of the MC simulation when using it in this work.

Table 1. Comparison of the relevant conflict probability computation methods [78], [80], [81], [83], [84].

Methods	Analytical approximations	MC simulation	Gridding methods	Markov chain approximations
Number of ships	Ship pair	Multi-ship	Multi-ship	Ship pair
Model allows non-stationary processes	No	Yes	Yes	No
Model allows non-Gaussian processes	No	Yes	Yes	Yes
Model allows dependency between variables	No	Yes	Yes	Yes
Type of conflict measure models	Circular	Any	Circular	Circular and elliptical
Computational costs	Very low	High	Very High	Low

The second challenge in probabilistic CD is to model the perturbations influencing the ship movements, i.e., the PDFs of future prediction trajectories. Commonly, the position and/or heading course prediction errors in the literature are assumed to be Gaussian (or approximated Gaussian) distributions with zero mean [15], [16], [85], [86] and their variances are expected to grow linearly or quadratically with time [15], [20], [77], [87]. However, these hypotheses may be problematic and need to be verified, to avoid any false conflict estimation and wrong conflict avoidance decision considerably. In addition, there is some probabilistic CD research that takes into account the trajectory uncertainty using the reachable sets [25], [70], [88]. Despite they can provide the reachable domain boundaries of all possible ship or aircraft movements, but cannot offer the specified probability distribution of the potential states in the reachable sets. AIS data, as a valuable source of information, is adopted for diverse purposes in the maritime traffic domain, such as maritime anomaly detection [89], marine traffic movement characterization [90], [91], trajectory pattern recognition [92], [93] and vessel behavior analysis and inference [94], [95]. Hence, it is promising to extract the ship spatiotemporal motion uncertain patterns from historical AIS data.

In this study, we propose a probabilistic CD approach to address the above challenges by developing both a two-stage MC simulation algorithm for efficient and accurate conflict probability computation and an AIS data-driven procedure for ship motion uncertainty pattern mining. As a result, the findings will facilitate the ship owners and

maritime authorities to detect collision risks in complex maritime traffic waters.

3 Methodology: A probabilistic conflict detection approach

To provide a quantitative basis for decision-making by the crew or Vessel Traffic Service (VTS) center in port waters with high traffic densities, a novel probabilistic CD approach is designed in which the uncertainty caused by environmental disturbances, mechanical as well as human factors is considered. It is designed to identify the target ships with potential collision risk and quantify the occurrence probability of conflict by incorporating the spatial-temporal movement uncertainty of multiple ships. This CD approach is characterized by the following blocks, as shown in Fig. 2. First, the basic concept of ship conflict and its criticality measure model are introduced to assess how safe the current ship encounter is. Secondly, the ship position in the look-ahead time horizon is predicted by incorporating the information on the ship navigation plan and the characteristics of the disturbances affecting the ship motion. Various sources of uncertainty may cause deviations of the prediction position. Thus, we extract the PDFs of trajectory uncertainty components using historical AIS data to model the future trajectory probability distribution at each instantaneous time. Thirdly, the conflict probabilities in multi-ship encounters are computed based on the prediction position distributions to decide whether to issue an alert or not. For each ship, the minimal passing distances with its nearby ships are first computed to identify the target ships with potential collision risk, then the conflict probabilities with these ships are estimated using a fast improved MC algorithm implementation.

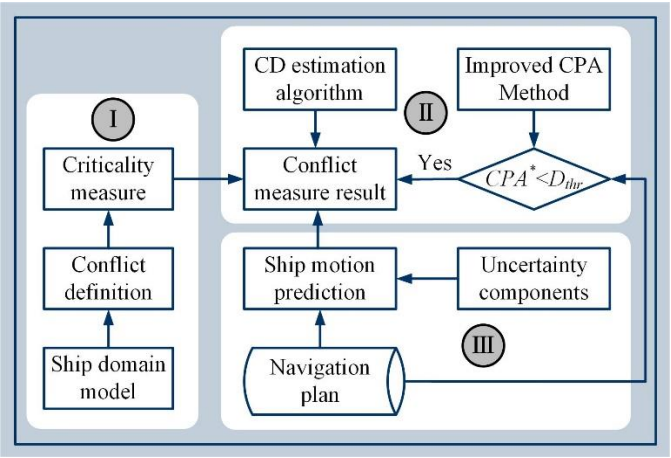


Fig. 2. Framework for conflict detection of ship encounters.

3.1 Conflict definition and its criticality measure

A conflict occurs when the trajectories of two vessels are predicted to violate a given set of prescribed separation distances. In this study, vessel conflict is defined based on the ship domain model. An example of conflict identification is illustrated in Fig. 3. In this figure, ships *A* and *B* are considered to be in conflict if the following

formula is held in the near future time.

$$Dist_{AB}(t) \le SD_A(t) + SD_B(t) \quad (1)$$

where SD_A and SD_B are the distances from each ship center to the boundaries of their ship domain area, $Dist_{AB}$ is the distance between the own ship and target ship. The widely used ship domain model from [38] that is suitable for the restricted areas with high-density traffic is adopted, which is an ellipse with a long radius of 6 L (L is ship length) and a short radius of 1.6 L. In fact, the shapes and sizes of ship domains are heavily dependent on the study waters' traffic density and traffic rules. An alternative approach is to design a ship domain model based on particular water areas' AIS data mining [45], [50], to determine the relationship between such impact factors as ship attributes, navigational environment and human factors, and domain sizes. However, the main concern in this paper is to identify and quantify the conflict of a multi-ship encounter under the presence of ship motion uncertainty. It is the probability that the ship trajectories experience the violation of a set of the minimum allowed distances, thus causing a conflict.

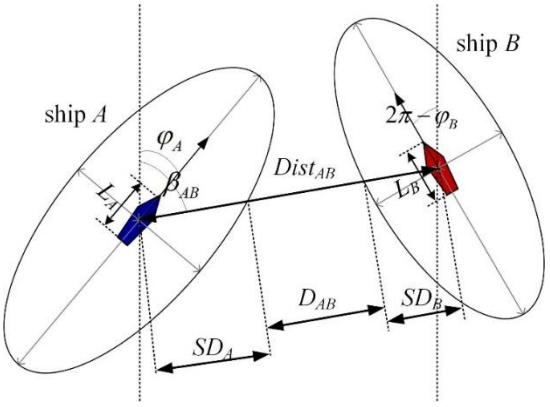


Fig. 3. Definition of ship conflict.

The instantaneous probability of a conflict at time t is given by the probability that the separation of the two ships is smaller than or equal to the prescribed separation distance, i.e. $Dist(t) \le SD_A + SD_B$, as follows:

$$PC(t) = \Pr[L(t) \le 0] = \int_{-\infty}^{0} f_{L(t)}(\rho) d\rho \quad (2)$$

where $f_{L(t)}$ represents the PDF of $Dist(t)$-SD_A-SD_B.

To characterize an appropriate supporting metric to measure the criticality of a conflict, we declare a conflict based on the maximum value of the conflict probabilities over a prediction horizon, as follows:

$$C(\gamma) := \max_{t \in [0,T]} PC(t) \quad (3)$$

where T is the predicted time horizon.

3.2 Model of the ship motion

CD is performed based on the predicted ship trajectories. According to the survey in [5], ship trajectory prediction

methods can be categorized into three groups: Physics-based, Maneuver-based and Interaction-aware prediction. In these methods, Interaction-aware prediction is considered to be the most accurate. This is due to the fact that this type of methods requires the exchange of planned trajectories among ships through communications, while each ship has better knowledge about its own intentions and trajectory information than other ships. Therefore, we assume that the planned trajectory information or navigation plans are obtained based on the interaction among ships [96], [97].

3.2.1 Vessel absolute motion

Typically, a ship navigation plan consists of a sequence of waypoints $WP_{i=1,2\ldots,n+1}$, which specifies a piecewise linear nominal trajectory. For the purpose of conflict estimation, we first compute the nominal ship trajectory with the assumption that each ship follows its navigation plan moving along the line connecting successive waypoints with a prescribed speed, then the position uncertainty is added to the nominal trajectory, from which the occurrence probability of conflict can be computed. As a result, the vessel motion model in this study is composed of the following three components: 1) a continuous dynamics describing the laws of physics of the vessel motion; 2) a discrete dynamics associated with the navigation plan; and 3) a stochastic component given by the ship motion uncertainty caused by environmental disturbances such as wind, waves and currents, as well as mechanical and human factors, etc.

Building on the above model, the predicted position of ship A in the future T moment can be expressed as follows:

$$\vec{S}_A(t_c+T)=\vec{S}_A(t_c)+\int_{t_c}^{t_c+T}\vec{V}_A(t)dt+R(\varphi_A(T))\cdot\vec{Q}_A(T) \quad (4)$$

where t_c is the current time instant; $\vec{S}_A(t_c)$ is the initial position of ship A; $\vec{V}_A(t)$ denotes the nominal speed of ship A at time t, which is a piecewise constant functions associated with the navigation plan; $R(\varphi_A(T))$ is the rotation matrix associated with the ship's nominal heading course $\varphi_A(T)$; and $\vec{Q}_A(T)=[Q_{A,x}(T);Q_{A,y}(T)]$ represents the uncertain components of ship prediction position. More details about Eq. (4) can be seen in Appendix A.

3.2.2 Vessel relative motion

As a result of the conflict occurrence probability being highly dependent on the relative motion between the encountering ships, the distance between them is first given as follows (from Eq. (4)).

$$Dist_{AB}(t)=\|\vec{S}_A(t)-\vec{S}_B(t)\| \quad (5)$$

In addition, the ship domain boundary relations between encountering ships also have a significant effect on the conflict criticality (see Fig. 3). Thus, the distance from the center of ship A to its ship domain boundary along the line between the positions of the two vessels is given by the following expression.

$$SD_A(t)=\left(\frac{1+\tan^2(\beta_{AB}(t)-\varphi_A^T(t))}{\frac{1}{R_{L,A}^{\ 2}}+\frac{\tan^2(\beta_{AB}(t)-\varphi_A^T(t))}{R_{S,A}^{\ 2}}}\right)^{1/2} \quad (6)$$

where $\beta_{AB}(t)$ represents the predicted relative course of the position of ship B over that of ship A at time t (see Fig. 3), $R_{L,A}$ and $R_{S,A}$ are the length of the semi-major axis and semi-minor axis of ship A's domain ellipse, and $\varphi_A^T(t)$ denotes the predicted heading course of ship A at time t. As the ship's course may vary slightly when sailing along the route derived from the navigation plan under the effect of various disturbances, its prediction uncertainty component has also been considered due to its significant impact on the length of SD_A. As a result, $\varphi_A^T(t)$ can be described as follows:

$$\varphi_A^T(t)=\varphi_A(t)+\alpha_A(t) \quad (7)$$

where $\alpha_A(t)$ represents the predicted course error component at time t, which will be introduced in detail in Section 3.2.3. In a similar way, the length of SD_B can also be obtained.

3.2.3 Extraction of ship position and course uncertainty patterns

One of the most important tasks in the ship motion modelling is to identify the distribution functions of uncertainty components influencing the ship motion (i.e. $Q_{A,x}$, $Q_{A,y}$ and α_A in this study), since the accuracy of the estimated $C(\gamma)$ relies on the predicted ship state distributions to a large extent. To tackle this issue, we design an AIS data-driven procedure to determine the PDFs of position and course uncertainty components. It consists of four steps: 1) Identification of trajectories' turning points, 2) Extraction of position and course prediction errors, 3) Correlation test of uncertain component datasets, and 4) PDF fitting of uncertain components.

To be more specific, the AIS data trajectories' turning points are first identified as the waypoints of ship navigation plans in terms of Douglas–Peucker (DP) algorithm [98]. This algorithm can compress line data in one trajectory by splitting them recursively to retain the important trajectory positions. Due to its great performance in running speed and accuracy, it has been widely used in vessel trajectory compression [99], [100]. Therefore, we adopt it for simplifying the trajectories and identifying the ships' turn points. The details about the DP algorithm design are found in [101].

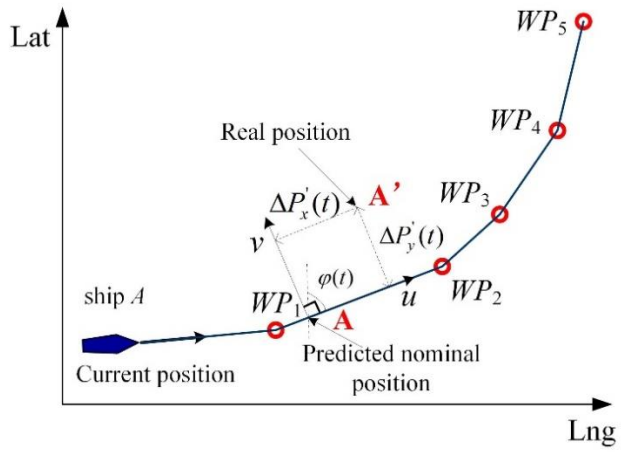


Fig. 4. Illustration of computation of position and course prediction errors.

On the above basis, we can predict each ship's future nominal trajectory based on their current state and the identified turn points. By computing the differences between the nominal prediction trajectory and the real trajectory from historical AIS data, the position and course prediction errors at each time moment over the predicted time horizon are extracted. An example of error computation is given in Fig. 4. In this figure, points *A* and *A'* represent the predicted nominal position and the real position at time *t*, respectively. Considering a route-fitted coordinate system with *u* aligned with the ship's nominal sailing direction and *v* perpendicular to it, the ship's prediction position errors in its vertical and horizontal directions can be computed, as follows:

$$\Delta P_x^{'}(t) = \sin\varphi(t)\cdot\Delta P_x(t) + \cos\varphi(t)\cdot\Delta P_y(t) \quad (8)$$

$$\Delta P_y^{'}(t) = -\cos\varphi(t)\cdot\Delta P_x(t) + \sin\varphi(t)\cdot\Delta P_y(t) \quad (9)$$

where $\varphi(t)$ is the ship's predicted nominal course in time *t*, $\Delta P_x(t)$ and $\Delta P_y(t)$ represent the predicted longitudinal and lateral position errors on the original geographic coordinate system. As for the course error, it can be easily extracted based on the difference between the nominal prediction course and the real course. In this way, we can collect the datasets of the ships' nominal prediction position and course errors for every minute over the prediction time horizon. Since the correlations between these error datasets have an important impact on the evaluation accuracy when computing the probability of a conflict, the Pearson correlation coefficient [102] is adopted to measure the dependence between each pair of datasets with the same time, before performing dataset PDF fitting.

Finally, the Kernel Density Estimation (KDE), a nor-paramedic estimation method [103], is adopted to identify the PDFs of these datasets, through the following equation:

$$g(x) = \frac{1}{K'}\sum_{i=1}^{K'}\phi_h(x - x_i) = \frac{1}{K'h}\sum_{i=1}^{K'}\phi_h\left(\frac{x - x_i}{h}\right) \quad (10)$$

where ϕ_h is a kernel function with window bandwidth h that satisfies $\phi_h(x)>0$ and $\int_R \phi_h(x)dx = 1$, K' denotes the number of elements in the dataset to be investigated within the bandwidth h. In this study, we employ the Gaussian kernel to determine the PDFs.

It should be noted that the $Dist_{AB}$, SD_A and SD_B in Inequality (1) are functions of the position and heading course prediction error components ($Q_{A,x}$, $Q_{A,y}$, $Q_{B,x}$, $Q_{B,y}$, α_A and α_B), so whether Inequality (1) is held is a probabilistic event and its occurrence probability needs to be determined based on probability computation methods.

3.3 Conflict probability estimation

In reality, the officer on watch (OOW) needs to detect target ships with potential collision risk from a huge number of sailing ships within given busy water, before conducting ship conflict or collision risk estimation. For that, we first adopt an improved CPA method to extract target ships spatially positioned close to the own ship in the near future, then a two-stage MC simulation algorithm is presented to estimate the $C(\gamma)$ level in multi-ship encounters.

3.3.1 Identification of target ships with potential collision risk

The traditional way to calculate the minimum passing distance of two encountering ships is based on the CPA method. However, as mentioned above, this approach is used with the assumption that the ships are sailing linearly without changes in heading and speeds. To tackle this issue, we adjust the traditional CPA method to make it adaptive to non-linear ship motion cases. Appendix B presents the equations for the calculation of minimum passing distance between ships. As a preliminary step of conflict probability estimation, the identification of potential conflict ships is undertaken based on the nominal prediction trajectories derived from the navigation plans. After obtaining the values of minimum passing distance, we can preliminarily identify whether there are potential collision risks between ships in the near future.

3.3.2 Computation of conflict criticality in multi-ship encounters

One of the biggest obstacles in the implementation of the CD approach is the computation of the probability of conflict occurrence, since there exists no derived analytical solution for $PC(t)$ in Eq. (2). The MC simulation is used for a solution [84]. Applications of a Direct Monte Carlo (DMC) simulation are often computationally intensive, and hence speed-up improvement for DMC is essential for certain online applications.

For a typical MC simulation, it consists of two loops, one for sampling iterations and another for trajectory propagation [78]. In the sampling iteration loop, it generates N groups of samples of random variables in terms of their prescribed PDFs, and each group of these samples is then inserted into the stochastic model to find a deterministic solution. By using the ensemble of the deterministic solutions, one can finally obtain an approximate estimated value. As the value of N in the sampling iteration loop defines the accuracy of the estimated solution, we

need to look for ways to improve the computation efficiency of DMC from the trajectory propagation loop. In general, the values of conflict probabilities of encountering ships over a finite look-ahead time horizon tend to exhibit a sharp spike at some time instant, whereas are relatively small elsewhere. In view of the fact that only the maximum value of the conflict probabilities over the prediction horizon requires to be estimated exactly, we can roughly extract the time points with higher conflict probabilities before executing a large number of iterations. Based on this principle, a two-stage MC simulation algorithm is developed to efficiently estimate the conflict criticality, in which the quantitative bounds on the approximation accuracy is also provided.

To decide how many iterations (N) are adequate to guarantee a desired accuracy of the estimation, Hoeffding's inequality [83] that describes the relation between N and the estimated accuracy is first given as follows:

$$N \geq \frac{\log(2/\delta)(b-a)^2}{2\varepsilon^2} \quad (11)$$

where ε denotes the accuracy, 1-δ represents the confidence, a and b represent the lower and upper bounds of estimated values, which are 0 and 1, respectively. This inequality indicates that for finite N, one can ensure a certain accuracy ε for the estimator with confidence 1-δ. In this study, the accuracy is designed as 1%, in order to provide an accurate conflict warning for practical applications. It requires a total of 15000 iterations to achieve this accuracy with a confidence $1-2\times10^{-3}$.

Algorithm 1. Two-stage MC simulation algorithm.

Require: PDFs of random variables in different time moments $Q_x^{(t)}$, $Q_y^{(t)}$ and $\alpha^{(t)}$, number of iterations in the first stage N_{MC1}, number of iterations in the second stage N_{MC2}, prediction time horizon T, number of ships N.

1: // Extract time point with higher conflict probabilities for ship A
2: **For** t = 1, 2,…, T **do**
3: Generate random sample vectors for each ship~$\{Q_{l,x}^{(t)}, Q_{l,y}^{(t)}$ and $\alpha_l^{(t)}$: $l = A, B, \ldots, N\}$ // length of sample vectors N_{MC1}
4: Compute the values of vectors SD_A, SD_B, …, SD_N; $Dist_{AB}$, …, $Dist_{AN}$ based on generated random sample vectors
5: c = count($SD_A + SD_B > Dist_{AB} \| \ldots \| SD_A + SD_N > Dist_{AN}$)
6: $PC_1(t) = c/N_{MC1}$
7: **End for**
8: Find and rank the time points whose probability values are larger than $\max\{PC_1(t)\} - 2\times[\log(2/\delta)/(2\times N_{MC1})]^{0.5}$, and then store at most the first two time points into T_2
9: // Execute a large number of iterations for time points in T_2
10: **For** $t \in T_2$ **do**
11: Repeat step 3-6 to obtain $PC_2(t)$ // length of sample vectors N_{MC2}
12: $PC(t) = (PC_1(t)\times N_{MC1} + PC_2(t)\times N_{MC2})/(N_{MC1}+N_{MC2})$
13: **End for**
14: $C(\gamma)=\max\{PC(t)\}$ $t \in T_2$

Output: $C(\gamma)$

Algorithm 1 provides a detailed description of the proposed two-stage MC simulation algorithm. At the first stage, the conflict probability at each time instant over the prediction time horizon is computed roughly with a relatively

small number of samples (e.g. 1000 iterations). Then the time points which may have the maximum probability of conflict are extracted and ranked in combination with the quantitative bounds on the approximation error obtained from Inequality (11). In this process, at most the first two ranked time points are retained for further accurate conflict probability computation when facing the cases where many time points are extracted, since retaining too many time points will increase the computational burden in the second stage. It should be pointed out that the cases that are required to extract too many time points usually occur when the estimated $C(\gamma)$ is small. In practice, if the estimated maximum conflict probability in the first stage is far below the threshold of conflict warning, it is not necessary to continue the conflict probability computation in the second stage, so as to reduce the waste of computational resources. In addition, the bounds derived from Inequality (11) are generally conservative [83], which means the real number of time points that may have the maximum conflict probability is less than that we extract. Therefore, it is acceptable to retain a small number of time points for the second stage. On the above basis, we can further obtain a more accurate $C(\gamma)$ with a large number of iterations in the second stage. It is noteworthy that this approach is generic and can be adaptive to multiple ships (see lines 3-6 in Algorithm 1).

4 Applications and case study results

4.1 Case description and data

To demonstrate the effectiveness and applicability of the proposed CD approach as well as extract the ship position and course uncertainty patterns, a total of 24842542 AIS data records from the Ningbo-Zhoushan port waters during one month period from October. 1, 2018 to October. 31, 2018 are collected. This water area is located between longitudes 121°52E-122°22E and latitudes 29°43N-30°02N (see Fig. 5). It is located in the world largest port in terms of cargo throughput, exhibiting frequent multi-ship encounters due to its high traffic densities over space and time. In addition, the ships sailing from and to the port turn frequently constrained by the port geometry. Therefore, it is chosen to be the representative and experimental port water for testing the proposed CD approach.

Although the ship information provided by AIS data is highly reliable, the collected data are not immune to errors occurring due to data transmission or other causes. To remove these errors and improve the data quality for experimental analysis, AIS data pre-processing need to be implemented to cleanse the data. In this study, AIS data is composed of the following format: Maritime Mobile Service Identity (MMSI) number, time, longitude, latitude, speed over ground (SOG), course over ground (COG), type, length and breadth. We take a data pre-processing procedure including noise elimination in each attribute [104], trajectory extraction and separation, confirmation of trajectory consistency [105] and data interpolation [106] to ensure the effectiveness and reliability of data.

Fig. 5. Study area.

After completing the data pre-processing procedure, we select all the AIS messages of cargo ships and tankers as our experiment data source, including a total of 59989 sailing pattern trajectories of 8582 vessels. It should be noted that the trajectories containing less than 60 points or 20 min duration are excluded because they are small and cannot fully reflect the track features of interest.

4.2 Fitting distribution functions of ship position and course uncertainty patterns

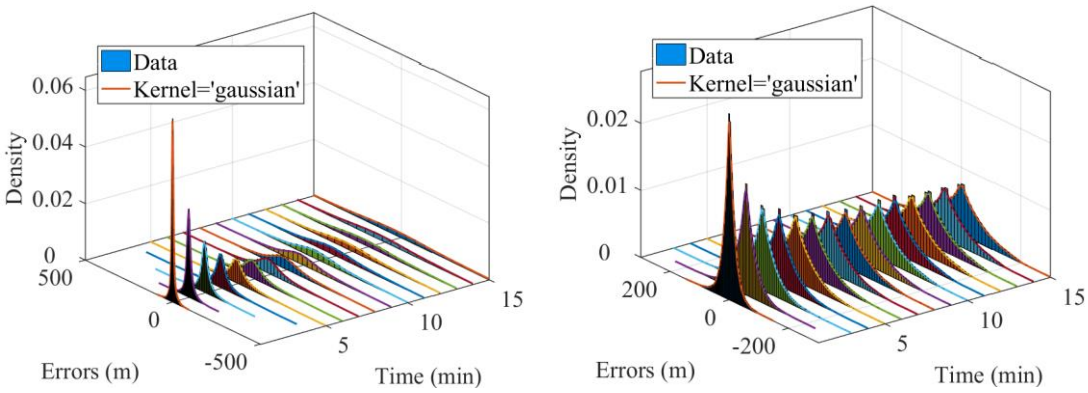
In this study, we collect the datasets of the ships' nominal prediction position and course errors for every minute over a 15 min horizon (the collision-warning time is set to be 15 min), by utilizing each ship's trajectory for error sampling. The Pearson correlation results show that except for the high correlation coefficients between the horizontal position errors and course errors in 1 and 2 minutes (larger than 0.4), the remaining dataset pairs are insignificantly correlated and satisfy the hypothesis of no correlation with a significance level of 5% [107]. For simplicity, all error datasets are assumed to be statistically independent. In fact, this assumption is feasible for practical application since the extracted prediction errors in the first two minutes are smaller compared with those with prediction times larger than 2 min, which are relatively less influential to the $C(\gamma)$ outputs.

To check whether the position or course error datasets agree with a normal distribution, the Jarque–Bera test [108] is selected for experimental testing. It is found that none of them fit a Gaussian distribution. That is to say, the hypothesis that the Gaussian distribution can be utilized to reflect the trajectory prediction uncertainty is inaccurate [15], [16], [18]–[20]. In addition, some typical density distribution functions such as *t* Location-Scale, Stable, Logistic, Extreme Value and Generalized Extreme Value are also considered. The Kolmogorov-Smirnov test [109] is used to measure how well these distributions fit the collected datasets. The results show that these distribution functions are still not appropriate because all *P*-values obtained using the Kolmogorov-Smirnov test are below 0.05. Consequently, the PDFs of these datasets are finally identified using the KDE method, because it has obvious superiority in fitting any

distribution shape [45]. The corresponding fitting results are illustrated in Fig 6.

Fig. 6 (a)-(c) depict the error datasets' PDF fitting curves along with their normalized bar charts for every minute over a 15 min look-ahead horizon. All these curves are approximately symmetric, with sharp turning points occurring near 0 and progressively descending to both sides. From Fig. 6 (a), it is noticeable that the curves rapidly become lower and wider with time, indicating that the along-track prediction uncertainty grows significantly in time. In contrast, the cross-track error PDF curves exhibit a significantly different trend. According to Fig. 6 (b), the curves change over time considerably in the initial stage but basically remain stable for the rest of time horizon. As for the fitting curves of course errors (see Fig. 6 (c)), there are no significant changes with time.

To further identify the change rates of these error datasets, the standard deviations for different uncertainty components over time are presented in Fig. 6 (d). It can be found that the variations of standard deviations with time are basically consistent with those of fitting curves. Both the along-track and course error standard deviations grow linearly with time, in which the latter increase with a slower speed, while the cross-track error standard deviations increase fast initially but remain a smaller growth rate for prediction times larger than 3 min. In addition, it can be observed that the standard deviation of along-track errors is significantly larger than that of cross-track errors, especially as time goes on, reflecting the fact that the prediction trajectories in the vertical direction have a higher level of uncertainty. By embedding the PDFs obtained above into the CD approach, we can achieve the real-time identification of the $C(\gamma)$ levels.



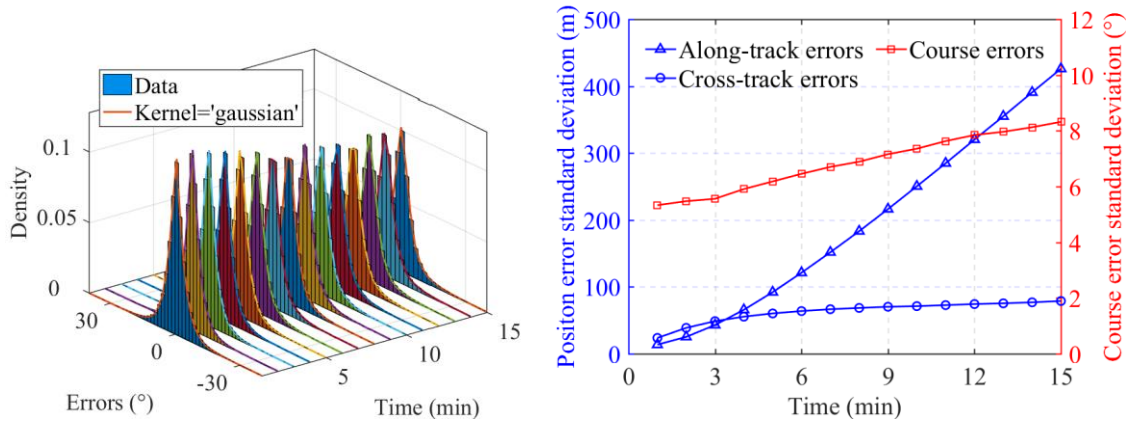


Fig. 6. (a) Time-dependent PDFs of along-track prediction errors; (b) time-dependent PDFs of cross-track prediction errors; (c) time-dependent PDFs of course prediction errors; (d) standard deviations of position and course errors over the prediction time horizon.

4.3 Experiments and results

In this subsection, several real experiments are performed to test and demonstrate the performance of the proposed probabilistic CD approach. We start by checking the accuracy and efficiency of the proposed two-stage MC through comparison with DMC. Then, two ship encounter scenarios including one ship-pair encounter and one multi-ship encounter derived from AIS data are introduced to explain how the proposed method is suitable for high traffic waters with complicated encounter situations. Applications of the proposed method on both real-time CD and off-line collision risk Hot-Spot identification are finally described.

4.3.1 Computational performance comparison

Fig. 7 provides the error statistics for the accuracy and computational costs computed by the DMC and two-stage MC (TSMC) with the different number of ships in encounter scenarios. For the computation of estimation errors of the two methods, the values obtained by performing the DMC simulation with 1,000,000 iterations are regarded as the true values. According to Fig. 7 (a), there is no significant difference in the Boxplot of error statistics of the two methods, and almost all sampled results based on the two methods have an error within 1%. Furthermore, in terms of the Root Mean Square (RMS) errors of the estimated accuracy (see Fig. 7 (b)), it can be found the proposed TSMC slightly underperform the DMC with the different number of encountering ships. However, in the term of the performance of running efficiency, the proposed method has a huge advantage. It is clear from Fig. 7 (b) that compared with the DMC which is computationally demanding, the computational costs of the proposed method are much lower and the advantage becomes more obvious with the increasing number of ships. That is, the proposed method requires much less computational costs to yield the same accuracy as the DMC. Therefore, the proposed

method can greatly enhance the computational efficiency while ensuring the accuracy of the approximate solution.

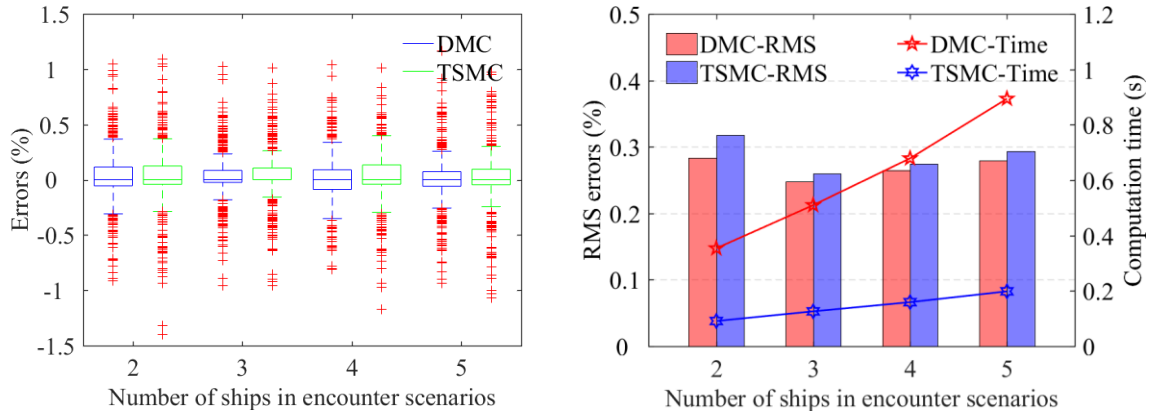


Fig. 7. Comparison between DMC and TSMC in accuracy and computation efficiency. (a) Boxplot of estimation error distributions for DMC (N=15000) and TSMC (N_{MC1}=1000, N_{MC2}=14000); (b) RMS errors and computational costs for DMC and TSMC.

4.3.2 CD approach test based on ship encounter scenarios

To test the performance of the proposed CD approach under the condition that the encountering ships have changeable spatiotemporal motion behaviors, a ship-pair encounter scenario derived from the historical AIS data is presented (see Fig. 8). The lines in Fig. 8 (a) are the trajectories of the ship-pair involved in the encounter, where 'x' marks the ships' starting locations and '△' their final locations. It can be seen that one ship basically sails linearly, while another one has a turning behavior during the encounter.

Fig. 8 (b) displays the risk variable evolutions over time for the encounter scenario. In the figure, the $C(\gamma)$ has a negative relation with the minimum passing distance (CPA^*) over a 15 min look-ahead time horizon (be in line with the setting of prediction time in the CD approach) computed by the improved CPA method, which coincides with common sense that a smaller minimal passing distance corresponds to a higher conflict probability. However, when comparing the $C(\gamma)$ with the minimum passing distance (CPA) computed by the original CPA technique, we cannot get similar results. It is found that CPA experiences two troughs due to the turning action of ship A, which may confuse ship owners in identifying collision dangers. Obviously, the original CPA technique becomes ineffective for this encounter scenario. This can be attributed to the fact that the CPA technique is performed with the assumption that ships are sailing linearly without changes in heading and speeds. As a result, all CPA-based collision risk assessment methods may provide false collision alerts, which hinders their applications in highly dynamic traffic situations.

In addition, one interesting phenomenon that must be mentioned is that the $C(\gamma)$ has dropped to 0 before the distance

between the ship-pair reaches the minimum. This is due to the fact that the ship conflict does not happen at the minimal distance, and their subsequent movements follow a diverging trend. However, when the two ships are approaching, the $C(\gamma)$ shows an upward trend and reaches a high level because of the potential conflicts caused by the uncertainty inherent to the ship spatiotemporal movements. Consequently, the proposed CD approach can detect the potential conflicts in advance by taking into account both the dynamic and uncertain characteristics of ship motion, thus providing exact and timely collision warning.

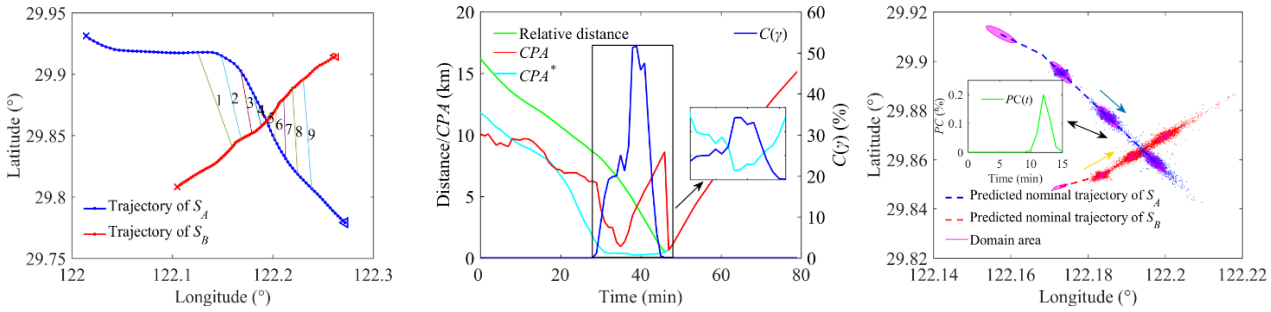


Fig. 8. A ship-pair encounter scenario. (a) Trajectories of ship pair; (b) risk variable evolutions of ship-pair encounter; (c) illustration of prediction trajectories and *PC* of ship-pair at time t=33 min.

Fig. 8 (c) provides an example to demonstrate how to obtain the $C(\gamma)$ under the presence of ship motion uncertainty. In the figure, the dash lines represent the nominal prediction trajectories over the prediction horizon, the pink ellipses are the ship domain areas in terms of the nominal prediction positions at prediction time 0, 5, 10 and 15 min, and the scattered points represent the potential position distributions at the corresponding prediction time due to ship spatiotemporal movement uncertainty. By incorporating the potential position distributions into the two-stage MC simulation algorithm, the *PC* at each prediction time slice can be computed (see the subgraph in Fig. 8 (c)) and the corresponding $C(\gamma)$ can be finally obtained.

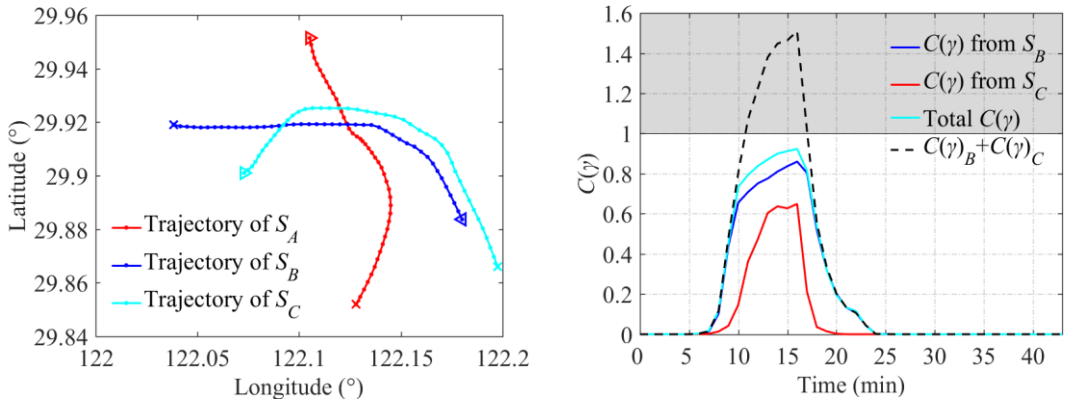


Fig. 9. A multi-ship encounter scenario. (a) Trajectories of multiple ships; (b) $C(\gamma)$ in the multi-ship encounter scenario.

To further validate the proposed CD approach in multi-ship encountering cases, a three-vessel encountering scenario derived from the AIS data is selected for experiments. The trajectories and the $C(\gamma)$ levels of the scenario are plotted in Fig. 9 (a) and (b). In Fig. 9 (b), the cyan curve represents the total $C(\gamma)$ of ship A in the multi-ship encounter, while the blue curve and red curve represent the $C(\gamma)$ of ship A with ship B and ship C, respectively. It can be observed that the total $C(\gamma)$ of ship A is higher than that with any single target ship. This finding meets the general knowledge about collision risk, i.e. a ship involved in multi-ship encounters usually faces greater risks than that involved in ship-pair encounters. Moreover, the total $C(\gamma)$ of ship A is not the sum of the $C(\gamma)$ with each single target ship, which can be verified by comparing the cyan line with the dotted line. In fact, the difference between the above two lines represents the probability of all encountering ships involved in conflicts. Therefore, the proposed CD approach not only can detect the own ship' conflict probability in multi-ship encounters, but also provide the occurrence probability of the multi-ship conflict.

As stated before, collision risk estimation that is adaptive to multiple ships is very essential for high traffic water areas. Here we statistically compare the own ships' total $C(\gamma)$ with their $C(\gamma)$ from the most dangerous target ships in multi-ship encounters to prove it (see Fig .10 (a)). In the statistics, we extract the multi-ship encounters every half an hour (48 times per day) based on our one month of AIS data, where only the target ships that have a $C(\gamma)$ level larger than 0.1 with the own ships are counted. From Fig .10 (a), the means of own ships' total $C(\gamma)$ in multi-ship encounters are significantly larger than those only from the most dangerous target ships and the more number of ships encounter, the more obvious the differences are. This indicates that focusing on the conflicts between ship-pairs is often not adequate since it has difficulty in reflecting the actual dangerous situations caused by multiple ships.

To further illustrate the high frequency of the multi-ship encounters, the total amount of multi-ship encounter conflicts and its compositions are presented in Fig .10 (b). According to the figure, a total of 11957 multi-ship encounters is counted and the three-ship encounters account for the largest percentage with 57%. As we extract the multi-ship encounters twice per hour, it implies that around 8.04 (11957/31/24/2) multi-ship encounter conflicts can be examined at each time moment. Considering both the severe conflict levels of multi-ship encounters and their high frequency, the focus should be placed on the CD of multiple ships rather than between ship pairs. As a result, our proposed approach can be an attractive option for the CD in waters with high traffic intensities.

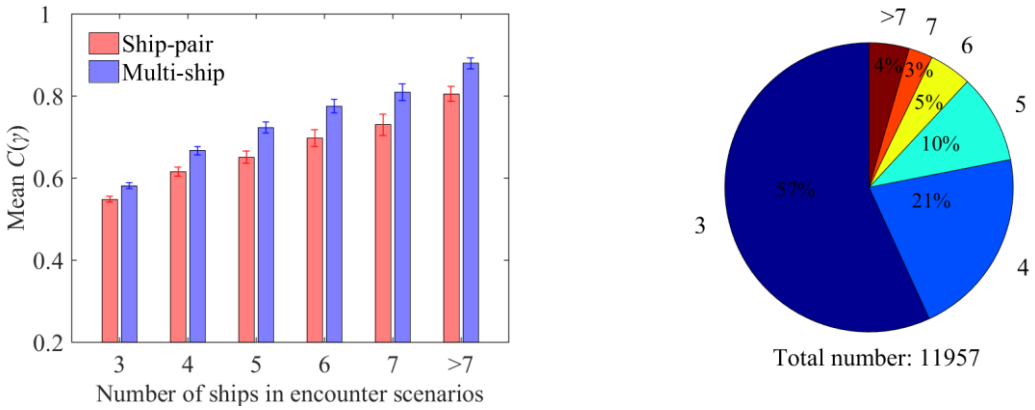
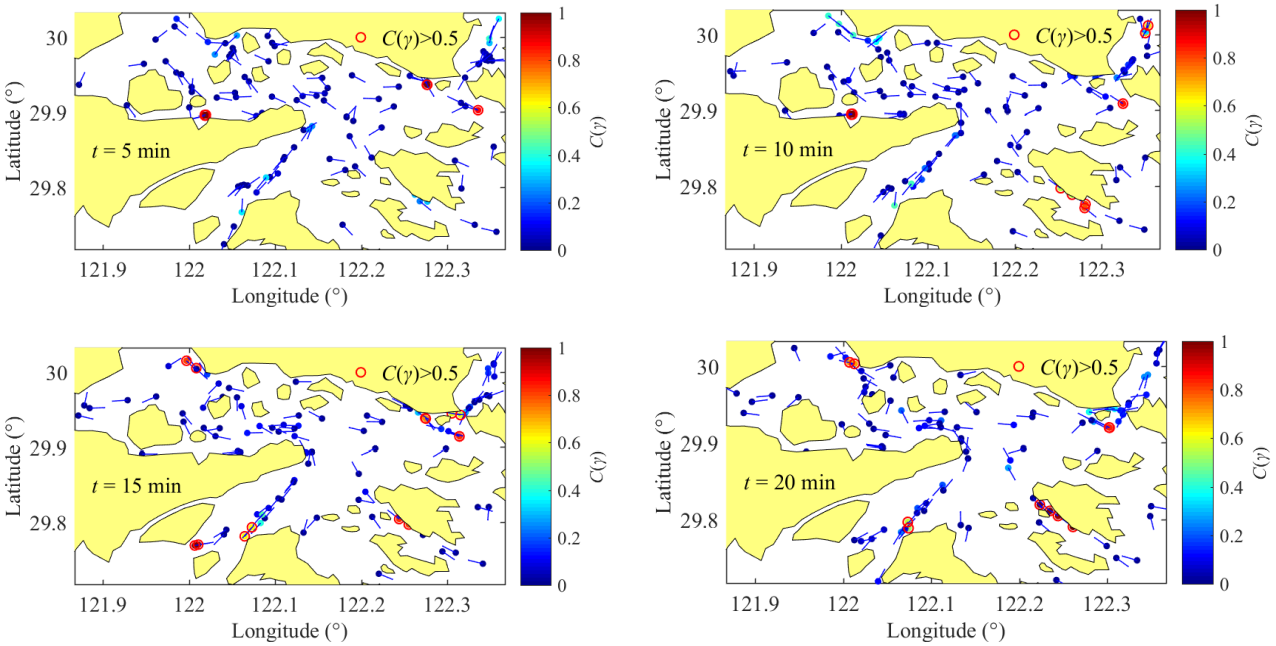


Fig. 10 (a) Comparison between the means of own ships' total $C(\gamma)$ and those with the most dangerous target ships in multi-ship encounter conflicts (the error bars represent the 95% confidence intervals of the means); (b) proportions of different number of ships involved in multi-ship encounter conflicts.

4.3.3 Application of the proposed probabilistic CD approach

From the practical viewpoint of application, the proposed CD approach has the potential to be applied for both real-time conflict estimation and off-line identification of conflict distribution characteristics. Thus, we demonstrate its online application in risk estimation through a case study. Then, the spatiotemporal features of the conflicts are explored in terms of the conflict distributions in space and time.



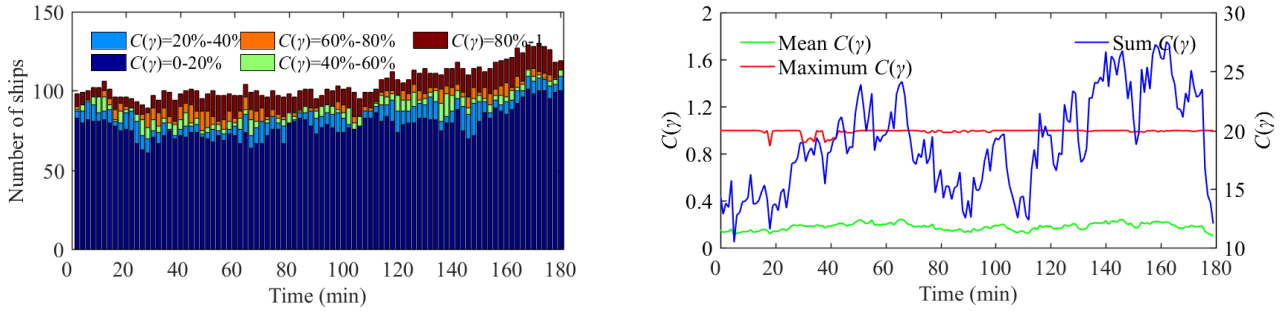


Fig. 11. Illustration of $C(\gamma)$ evolution of ship traffic for 3 hours in the study area. (a) ship traffic spatial distribution and their $C(\gamma)$ levels at t=5 min; (b) ship traffic spatial distribution and their $C(\gamma)$ levels at t=10 min; (c) ship traffic spatial distribution and their $C(\gamma)$ levels at t=15 min; (d) ship traffic spatial distribution and their $C(\gamma)$ levels at t=20 min; (e) compositions of ship traffic with different $C(\gamma)$ levels over time; (f) mean, maximum and sum $C(\gamma)$ curves of ship traffic over time.

Fig. 11 provides an example of the conflict evolution of ship traffic over time within the study area. In Fig. 11 (a)-(d), the points show the ships' positions, the blue lines represent the ships' heading course, and the points' colors display the ships' real-time $C(\gamma)$ levels. From these figures, it can easily find which ships will face high conflict probabilities in the near future (these ships with $C(\gamma)$ levels larger than 0.5 are circled with red circles), thereby providing ship owners with early warnings of potential collisions. Fig .11 (e) further depicts the compositions of ship traffic involved in different severity levels of conflicts. According to the figure, it can be observed that the number of ships generally shows an upward trend with time and we can also easily see how many ships are involved in high severity levels of conflicts at different time moments. Thus, from the perspective of maritime safety authorities, the proposed CD approach can assist them in monitoring and offering hazard warnings for high collision risk ships as well as facilitate them to implement risk mitigation measures timely.

In addition, we also investigate the $C(\gamma)$ evolution of ship traffic in terms of their mean, maximum and sum. It is seen from Fig .11 (f) that both the mean and maximum $C(\gamma)$ curves fluctuate slightly over time, and the latter basically remains stable at 1, implying that dangerous ship encounters occur at each time moment. In contrast, the sum $C(\gamma)$ changes considerably as time goes on, primarily because the amount of ships at different time varies greatly. By taking advantage of the three measures' ability to reflect the individual or total ship encountering conflicts, the traffic management center could better understand the real-time collision risk and traffic complexity comprehensively, so as to improve their working ability when facing dangerous traffic situations caused by high traffic intensities.

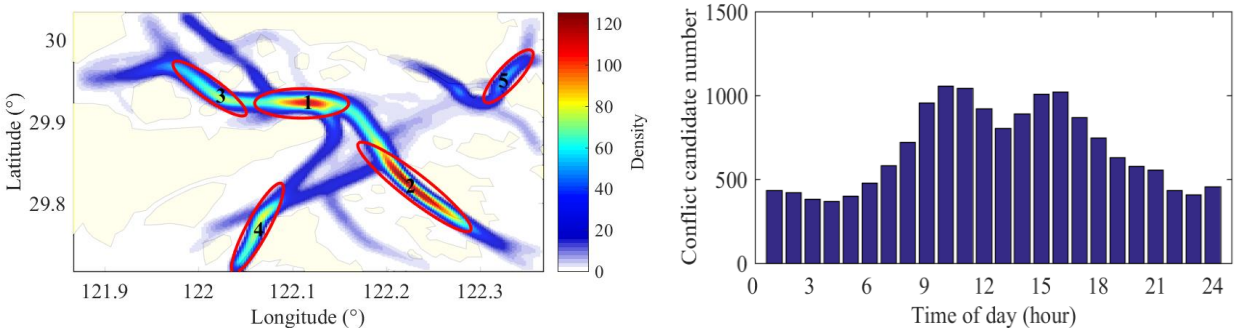


Fig. 12. (a) Visualization of spatial distribution of ship conflicts using KDE method; (b) Temporal distribution of ship conflicts (The horizontal axis with label 3 corresponds to 02:00-02:59).

Fig. 12 illustrates the spatial and temporal distributions of ship conflict candidates within the study area. Similar to the statistics in Fig. 10, we capture and retain the ship conflict candidates whose real-time $C(\gamma)$ levels are larger than 0.5 twice per hour. As shown in Fig. 12 (a), it can be easily seen that there are several conflict hotspot areas that are marked with red ellipses and with labels 1-5. Hotspots 1-3 belong to the main route that links the ship traffic between hub areas in the port and the outside waters, and consequently, having the largest traffic density and experiencing a high frequency of dangerous ship encounters.

In particular, hotspots 1 and 2 are associated with a higher frequency of ship conflicts compared with the other three areas. The higher conflict frequency for hotspot 1 can be explained by the fact that the ship traffic coming from other areas merges together in this area, increasing the frequency of multi-ship encounters. Besides, the frequently turning maneuvers of ships caused by the geometry constraint within this area may also contribute to this result, since the dynamic ship movements increase the difficulty of situational awareness of ship owners. In practice, hotspot 1 was an official precautionary area released by the Ningbo-zhoushan VTS center, displaying the effectiveness of the proposed CD approach on the identification of high collision risk areas to a certain extent. For hotspots 2, it is located in Xizhimen waterways, which is the main channel for the majority of large-scale ships to enter and exit the port but has narrower navigable width. Consequently, it becomes one of the riskiest areas for ship collision. The other two small hotspots (marked with number 4 and 5) are located in Fodu and Zhujiazui fairways. One possible reason for a little higher frequency of dangerous ship encounters in the two regions might also be that their narrow traffic widths result in the reduced minimal passing distances among encountering ships, thus producing lots of conflicts with high levels of severity.

It should be noted that compared with the existing conflict visualization studies [10], [11], [54] that identified the high collision risk areas based on the current state of ship-pairs, the proposed measures foresee the ships' potential conflicts in multi-ship encounters at present and in the near future time, thereby providing a different way to identify

high-risk areas.

In addition, Fig. 12 (b) clearly shows the number of conflict candidates against every hour of a day. A higher frequency of conflict candidates can be found during 08:00-12:00 and 13:00-17:00. This is consistent with the actual traffic situation in the regions, because a higher ship traffic density occurs during the daytime. Based on the identification results of conflict distributions in space and time, both ship owners and maritime authorities can gain valuable understanding about when and where to enhance situational awareness during vessel movements.

Our application analysis of real cases facilitates the validation of the proposed CD approach to estimate potential conflicts and identify areas with high collision risks. Therefore, its outcomes provide detailed insights on how to determine and implement appropriate conflict resolution strategies. For example, we can insert the proposed CD approach into the optimal control algorithms to ensure the resolution of potential conflicts in a complicated dynamic situation. The proposed approach can be widely used in collision risk monitoring and control.

4.4 Limitations and further improvements

Despite the advantages mentioned above, the proposed method has still revealed some limitations. Further studies are required to improve the research from the following aspects:

1. The ship trajectory prediction method needs to be further extended. In this study, the proposed approach highly depends on the exchange of navigation plans among ships through communications. However, in reality, non-cooperative ships are unlikely to share their navigation plans. Some manoeuver-based prediction techniques could be further incorporated into the model to support conflict evaluation to address the information non-disclosure issue among non-cooperative ships.
2. The proposed method emphasizes on the estimation of occurrence probability of a conflict. However, it may be inadequate to comprehensively assess the navigational risk, since the potential consequence is not explicitly considered in the conflict probability. In fact, there are a large number of possible accident scenarios with distinct occurrence probability and consequence once being involved in collisions. Therefore, an improved model which takes into account both the occurrence probability and consequence could aid ship owners and maritime safety authorities to better understand the actual level of risk of the traffic situation.
3. Some influencing factors (e.g. vessel type and human behaviors) need to be further investigated to critically analyze the extent to which they have impacts on the conflict criticality.

5 Conclusions

Collision risk analysis modelling multi-ship encounters is critical for marine traffic safety management, particularly

in complicated traffic waters. In this study, we develop a probabilistic CD approach to investigate the influence of spatiotemporal movement uncertainty of multiple ships on potential collision risk. The proposed approach has several unique features: 1) Both the dynamic and uncertain features of multi-ship movements are taken into account, so as to be applicable to various complicated encountering scenarios; 2) The developed conflict probability computation algorithm is efficient, accurate and having the capability to combine with any other conflict measure models (e.g., diverse "ship domain" [39], [40], [44], [45], [50] and "synthetic index" [13], [62], [63], [110]–[112]) without the requirement of changing its fundamental structure; 3) The spatiotemporal-dependent patterns of ship motions correlated with actual collisions are extracted and integrated to support a robust estimate of the collision risk.

Several experiments are carried out using real AIS-based trajectory data in the Ningbo-Zhoushan Port to test the performance of the proposed CD approach. The results show that the proposed approach performs better than these CPA-based approaches in detecting collision risks timely and reliably under a dynamic and uncertain traffic situation and can address multi-vessel encounter scenarios. The application analysis also demonstrates its effectiveness and applicability in both real-time CD and off-line collision risk Hot-Spot identification. As a result, the proposed quantitative approach provides detailed insights for ship owners, officers/captains and port management agencies into collision risk evaluation as well as facilitates them to implement risk mitigation measures.

Further effort will be made to enable the potential collision risk in both cooperative and non-cooperative scenarios to be evaluated by integrating some manoeuver-based prediction techniques into the proposed approach. In addition, it would be insightful to explore the global collision risk or ship traffic complexity of multiple ships under the presence of ship spatiotemporal motion uncertainty, to assist surveillance operators in improving their cognitive abilities in dangerous traffic situations.

CRediT authorship contribution statement

Xuri Xin: Methodology, Formal analysis, Investigation, Validation, Resources, Writing - original draft, Writing - review & editing. **Kezhong Liu:** Conceptualization, Methodology, Resources, Investigation, Validation, Supervision, Writing - review & editing, Project administration, Funding acquisition. **Zaili Yang:** Conceptualization, Methodology, Formal analysis, Investigation, Resources, Writing - original draft, Writing - review & editing, Supervision, Project administration, Funding acquisition. **Jinfen Zhang:** Methodology, Validation, Writing - review & editing, Supervision. **Xiaolie Wu:** Investigation, Formal analysis, Validation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was funded by the National Natural Science Foundation of China (Grant No. 52031009 & 52071247), China Scholarship Council (No. 202006950039) and National Key Technologies Research & Development Program (2017YFE0118000). This work is also financially supported by a European Research Council project (TRUST CoG 2019 864742).

Appendix A. Calculation of vessel absolute motion

In Eq. (4), the first two terms represent the nominal prediction trajectory based on the sequence of waypoints derived from the navigation plan. To be more specific, assume that the navigation plan of ship A comprises n_A segments with n_A+1 waypoints, in which the first waypoint is the initial position where the prediction is made. According to the assumption that each ship moves following its navigation plan with a prescribed speed, the nominal navigation time taken in each segment, as shown in Fig. A1, can be obtained by the following expression:

$$t_{A,i} = \frac{\| \vec{P}_{A,d_i} - \vec{P}_{A,o_i} \|}{\| \vec{V}_A(t) \|} \quad i = 1, 2 \cdots n_A \quad \text{(A1)}$$

where $\vec{P}_{A,d_i}$ and $\vec{P}_{A,o_i}$ denote the coordinates of the origin and destination waypoints of segment i, respectively. Then the nominal prediction times at which ship A starts and ends at each segment can be recursively computed by

$$t_{A,o_i} = \sum_{j=1}^{i-1} t_{A,j} \quad i = 2, 3 \cdots n_A \quad \text{(A2)}$$

$$t_{A,d_i} = \sum_{j=1}^{i} t_{A,j} \quad i = 1, 2 \cdots n_A \quad \text{(A3)}$$

Based on Eqs. (A1)-(A3), the nominal prediction position at any specific time can be easily computed.

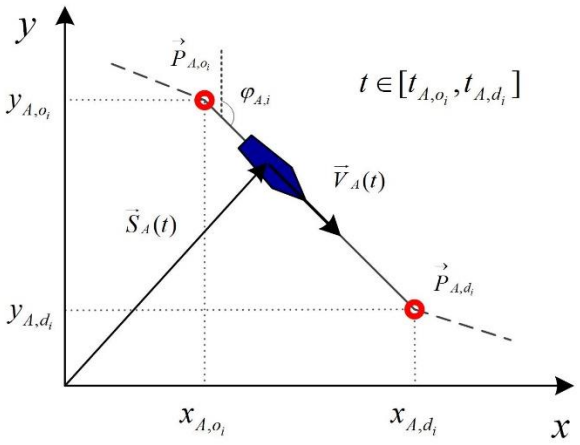


Fig. A1. Absolute motion of ship *A* in segment *i*.

For the last term in Eq. (4) that represents the ship prediction position errors added to the nominal trajectory, it is given in detail below.

$$R(\varphi_A(T)) \cdot \vec{Q}_A(T) = \begin{bmatrix} \sin(\varphi_A(T)) & -\cos(\varphi_A(T)) \\ \cos(\varphi_A(T)) & \sin(\varphi_A(T)) \end{bmatrix} \begin{bmatrix} Q_{A,x}(T) \\ Q_{A,y}(T) \end{bmatrix} \quad \text{(A4)}$$

where $Q_{A,x}(T)$ and $Q_{A,y}(T)$ refer to the heading and lateral ship position prediction error components, which are considered to be positive if they are toward to the front and left side, respectively. The nominal heading of ship *A* at time *T* can be computed based on the successive waypoint coordinates, as follows:

$$\varphi_A(T) = \arctan\left(\frac{y_{A,d_i} - y_{A,o_i}}{x_{A,d_i} - x_{A,o_i}} \right) \quad T \in [t_{A,o_i}, t_{A,d_i}] \quad \text{(A5)}$$

where (x_{A,o_i}, y_{A,o_i}) and (x_{A,d_i}, y_{A,d_i}) denote the positions of the origin and destination waypoints of segment *i*, with the exception of the case $x_{A,d_i} = x_{A,o_i}$, where $\varphi_A(T)=0/\pi$.

Appendix B. Calculation of minimum passing distance

Suppose the navigation plan trajectories of ships *A* and *B* comprise N_A and N_B segments, respectively. Then the nominal relative positions between the ship *A* sailing in segments $i = 1, 2, ..., N_A$ and the ship *B* sailing in segments $j = 1, 2, ..., N_B$ at time *t* can be formulated as follows:

$$\vec{S}^N_{AB,ij}(t) = \vec{S}^N_{A,i}(t) - \vec{S}^N_{B,j}(t) = \vec{S}^N_{0,AB,ij} + \vec{V}_{AB,ij} \cdot t \quad t \in [t_{A,o_i}, t_{A,d_i}] \cap [t_{B,o_j}, t_{B,d_j}] \quad \text{(B1)}$$

where $\vec{S}^N_{A,i}(t)$ and $\vec{S}^N_{B,j}(t)$ are the nominal prediction positions of ships *A* and *B* at time *t*, $\vec{V}_{AB,ij}$ represents the relative speed of ship *A* over ship *B* at time *t*, and $\vec{S}^N_{0,AB,ij}$ represents the derived relative initial positions of the two ships with the assumption of ship linear motion, which can be expressed as:

$$\vec{S}_{0,AB,ij}^{N} = \vec{P}_{A,o_i} - \vec{P}_{B,o_j} - (\vec{V}_{A,i} \cdot t_{A,o_i} - \vec{V}_{B,j} \cdot t_{B,o_j}) \quad \text{(B2)}$$

The distance between the two ships at time t is given by making use of Eq. (B1), as follows:

$$Dist_{AB,ij}^{N}(t) = \sqrt{\| \vec{S}_{0,AB,ij}^{N} \|^2 + 2\vec{S}_{0,AB,ij}^{N} \cdot \vec{V}_{AB,ij} \cdot t + \| \vec{V}_{AB,ij} \|^2 \cdot t^2} \quad \text{(B3)}$$

Since Eq. (B3) is a function with respect to t, the minimum distance between the ships for each pair of segments i and j is determined by t, as follows:

$$Dist_{\min,AB,ij}^{N} = \sqrt{\| \vec{S}_{0,AB,ij}^{N} \|^2 + 2\vec{S}_{0,AB,ij}^{N} \cdot \vec{V}_{AB,ij} \cdot t_{d_{\min},AB,ij} + \| \vec{V}_{AB,ij} \|^2 \cdot t_{d_{\min},AB,ij}^{2}} \quad \text{(B4)}$$

where $t_{d_{\min},AB,ij}$ denotes the time of the closest approach for the pair of segments i and j, which has the following three possibilities:

$$t_{d_{\min},AB,ij} = \begin{cases} \max\{t_{A,o_i}, t_{B,o_j}\} & t_{AB,ij}^{*} < \max\{t_{A,o_i}, t_{B,o_j}\} \\ -\vec{S}_{0,AB,ij}^{N} \cdot \vec{V}_{AB,ij} / V_{AB,ij}^{2} & t_{AB,ij}^{*} \in [t_{A,o_i}, t_{A,d_i}] \cap [t_{B,o_j}, t_{B,d_j}] \\ \min\{t_{A,d_i}, t_{B,d_j}\} & t_{AB,ij}^{*} > \min\{t_{A,d_i}, t_{B,d_j}\} \end{cases} \quad \text{(B5)}$$

After that, the minimum distance between the two ships over the prediction time horizon can be given as follows:

$$CPA_{AB}^{*} = \min_{i,j}\{Dist_{\min,AB,ij}^{N}\} \quad \text{(B6)}$$

References

[1] T. Chai, J. Weng, and X. De-qi, "Development of a quantitative risk assessment model for ship collisions in fairways," *Saf. Sci.*, vol. 91, pp. 71–83, 2017.

[2] P. Chen, Y. Huang, J. Mou, and P. van Gelder, "Probabilistic risk analysis for ship-ship collision: state-of-the-art," *Saf. Sci.*, vol. 117, pp. 108–122, 2019.

[3] L. Du, F. Goerlandt, and P. Kujala, "Review and analysis of methods for assessing maritime waterway risk based on non-accident critical events detected from AIS data," *Reliab. Eng. Syst. Saf.*, p. 106933, 2020.

[4] F. Goerlandt and J. Montewka, "Maritime transportation risk analysis: Review and analysis in light of some foundational issues," *Reliab. Eng. Syst. Saf.*, vol. 138, pp. 115–134, 2015.

[5] Y. Huang, L. Chen, P. Chen, R. R. Negenborn, and P. van Gelder, "Ship collision avoidance methods: State-of-the-art," *Saf. Sci.*, vol. 121, pp. 451–473, 2020.

[6] K. Kulkarni, F. Goerlandt, J. Li, O. V. Banda, and P. Kujala, "Preventing shipping accidents: Past, present, and future of waterway risk management with Baltic Sea focus," *Saf. Sci.*, vol. 129, p. 104798, 2020.

[7] E. Tu, G. Zhang, L. Rachmawati, E. Rajabally, and G.-B. Huang, "Exploiting AIS data for intelligent maritime navigation: A comprehensive survey from data to methodology," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 5, pp. 1559–1582, 2017.

[8] S. Li, Q. Meng, and X. Qu, "An Overview of Maritime Waterway Quantitative Risk Assessment Models," *Risk Anal.*, vol. 32, no. 3, pp. 496–512, 2012.

[9] A. Debnath and H. C. Chin, "Navigational traffic conflict technique: a proactive approach to quantitative measurement of collision risks in port waters," *J. Navig.*, vol. 63, no. 1, pp. 137–152, 2010.

[10] J. Weng, Q. Meng, and X. Qu, "Vessel Collision Frequency Estimation in the Singapore Strait," *J. Navig.*, vol.

65, no. 2, pp. 207–221, 2012.

[11] J. Weng and X. Shan, “Ship Collision Frequency Estimation in Port Fairways: A Case Study,” *J. Navig.*, vol. 68, no. 3, pp. 602–618, 2015.

[12] W. Zhang, F. Goerlandt, P. Kujala, and Y. Wang, “An advanced method for detecting possible near miss ship collisions from AIS data,” *Ocean Eng.*, vol. 124, pp. 141–156, 2016.

[13] W. Zhang, F. Goerlandt, J. Montewka, and P. Kujala, “A method for detecting possible near miss ship collisions from AIS data,” *Ocean Eng.*, vol. 107, pp. 60–69, 2015.

[14] W. Zhang, X. Feng, F. Goerlandt, and Q. Liu, “Towards a Convolutional Neural Network model for classifying regional ship collision risk levels for waterway risk analysis,” *Reliab. Eng. Syst. Saf.*, vol. 204, p. 107127, 2020.

[15] J. Park and J. Kim, “Predictive evaluation of ship collision risk using the concept of probability flow,” *IEEE J. Ocean. Eng.*, vol. 42, no. 4, pp. 836–845, 2016.

[16] H. Rong, A. P. Teixeira, and C. G. Soares, “Ship trajectory uncertainty prediction based on a Gaussian Process model,” *Ocean Eng.*, vol. 182, pp. 499–511, 2019.

[17] Z. Liu, Z. Wu, and Z. Zheng, “A cooperative game approach for assessing the collision risk in multi-vessel encountering,” *Ocean Eng.*, vol. 187, p. 106175, 2019.

[18] A. Lee, S. Weygandt, B. Schwartz, and J. Murphy, “Performance of trajectory models with wind uncertainty,” in *AIAA modeling and simulation technologies conference*, 2009, p. 5834.

[19] Y. Matsuno, T. Tsuchiya, J. Wei, I. Hwang, and N. Matayoshi, “Stochastic optimal control for aircraft conflict resolution under wind uncertainty,” *Aerosp. Sci. Technol.*, vol. 43, pp. 77–88, 2015.

[20] M. Prandini, J. Hu, J. Lygeros, and S. Sastry, “A probabilistic approach to aircraft conflict detection,” *IEEE Trans. Intell. Transp. Syst.*, vol. 1, no. 4, pp. 199–220, 2000.

[21] C. Chauvin, S. Lardjane, G. Morel, J.-P. Clostermann, and B. Langard, “Human and organisational factors in maritime accidents: Analysis of collisions at sea using the HFACS,” *Accid. Anal. Prev.*, vol. 59, pp. 26–37, 2013.

[22] R. J. Bye and A. L. Aalberg, “Maritime navigation accidents and risk indicators: An exploratory statistical analysis using AIS data and accident reports,” *Reliab. Eng. Syst. Saf.*, vol. 176, pp. 174–186, 2018.

[23] P. Kujala, M. Hänninen, T. Arola, and J. Ylitalo, “Analysis of the marine traffic safety in the Gulf of Finland,” *Reliab. Eng. Syst. Saf.*, vol. 94, no. 8, pp. 1349–1357, 2009.

[24] X. Qu, Q. Meng, and S. Li, “Analyses and implications of accidents in Singapore Strait,” *Transp. Res. Rec.*, vol. 2273, no. 1, pp. 106–111, 2012.

[25] H. Yu, Z. Fang, A. T. Murray, and G. Peng, “A Direction-Constrained Space-Time Prism-Based Approach for Quantifying Possible Multi-Ship Collision Risks,” *IEEE Trans. Intell. Transp. Syst.*, 2019.

[26] T. Macduff, “The probability of vessel collisions,” *Ocean Ind.*, vol. 9, no. 9, 1974.

[27] P. T. Pedersen, “Collision and grounding mechanics,” in *Danish Society of Naval Architects and Marine Engineers*, 1995, pp. 125–57.

[28] Y. T. Xi, Z. L. Yang, Q. G. Fang, W. J. Chen, and J. Wang, “A new hybrid approach to human error probability quantification–applications in maritime operations,” *Ocean Eng.*, vol. 138, pp. 45–54, 2017.

[29] M. R. Martins and M. C. Maturana, “Application of Bayesian Belief networks to the human reliability analysis of an oil tanker operation focusing on collision accidents,” *Reliab. Eng. Syst. Saf.*, vol. 110, pp. 89–109, 2013.

[30] S. Fan, E. Blanco-Davis, Z. Yang, J. Zhang, and X. Yan, “Incorporation of human factors into maritime accident analysis using a data-driven Bayesian network,” *Reliab. Eng. Syst. Saf.*, p. 107070, 2020.

[31] Q. Yu, K. Liu, C.-H. Chang, and Z. Yang, “Realising advanced risk assessment of vessel traffic flows near offshore wind farms,” *Reliab. Eng. Syst. Saf.*, vol. 203, p. 107086, 2020.

[32] F. Goerlandt and J. Montewka, “A framework for risk analysis of maritime transportation systems: a case study for oil spill from tankers in a ship–ship collision,” *Saf. Sci.*, vol. 76, pp. 42–66, 2015.

[33] J. Weng, G. Li, T. Chai, and D. Yang, “Evaluation of two-ship collision severity using ordered probit approaches,” *J. Navig.*, vol. 71, no. 4, pp. 822–836, 2018.

[34] P.-R. Lei, “Mining maritime traffic conflict trajectories from a massive AIS data,” *Knowl. Inf. Syst.*, vol. 62, no. 1, pp. 259–285, 2020.

[35] P. Chen, Y. Huang, J. Mou, and P. van Gelder, “Ship collision candidate detection method: A velocity obstacle approach,” *Ocean Eng.*, vol. 170, pp. 186–198, 2018.

[36] H. A. Nordkvist, “An advanced method for detecting exceptional vessel encounters in open waters from high resolution ais data.” NTNU, 2018.

[37] T. Watawana and A. Caldera, “Analyse Near Collision Situations of Ships Using Automatic Identification System Dataset,” in *2018 5th International Conference on Soft Computing & Machine Intelligence (ISCMI)*, 2018, pp. 155–162.

[38] Y. Fujii and K. Tanaka, “Traffic Capacity,” *J. Navig.*, vol. 24, no. 04, pp. 543–552, 1971.

[39] R. Szlapczynski and J. Szlapczynska, “An analysis of domain-based ship collision risk parameters,” *Ocean Eng.*, vol. 126, pp. 47–56, 2016.

[40] N. Wang, “A novel analytical framework for dynamic quaternion ship domains,” *J. Navig.*, vol. 66, no. 02, pp. 265–281, 2013.

[41] N. Wang, “An intelligent spatial collision risk based on the quaternion ship domain,” *J. Navig.*, vol. 63, no. 4, pp. 733–749, 2010.

[42] X. Qu, Q. Meng, and L. Suyi, “Ship collision risk assessment for the Singapore Strait,” *Accid. Anal. Prev.*, vol. 43, no. 6, pp. 2030–2036, 2011.

[43] F. Goerlandt and P. Kujala, “On the reliability and validity of ship–ship collision risk analysis in light of different perspectives on risk,” *Saf. Sci.*, vol. 62, pp. 348–365, 2014.

[44] N. Im and T. N. Luong, “Potential risk ship domain as a danger criterion for real-time ship collision risk evaluation,” *Ocean Eng.*, vol. 194, p. 106610, 2019.

[45] L. Zhang and Q. Meng, “Probabilistic ship domain with applications to ship collision risk assessment,” *Ocean Eng.*, vol. 186, p. 106130, 2019.

[46] J. Montewka, M. Gil, and K. Wróbel, “Discussion on the article by Zhang & Meng entitled “Probabilistic ship domain with applications to ship collision risk assessment” [Ocean Eng. 186 (2019) 106130],” *Ocean Eng.*, vol. 209, p. 107527, 2020.

[47] M. Gil, J. Montewka, P. Krata, T. Hinz, and S. Hirdaris, “Semi-dynamic ship domain in the encounter situation of two vessels,” in *Developments in the Collision and Grounding of Ships and Offshore Structures: Proceedings of the 8th International Conference on Collision and Grounding of Ships and Offshore Structures (ICCGS 2019), 21-23 October, 2019, Lisbon, Portugal*, 2019, p. 301.

[48] G. H. Dinh and N. K. Im, “The combination of analytical and statistical method to define polygonal ship domain and reflect human experiences in estimating dangerous area,” *Int. J. e-Navigation Marit. Econ.*, vol. 4, no. C, pp. 97–108, 2016.

[49] J. Liu, Z. Feng, Z. Li, M. Wang, and L. R. Wen, “Dynamic Ship Domain Models for Capacity Analysis of Restricted Water Channels,” *J. Navig.*, vol. 69, no. 3, pp. 481–503, 2016.

[50] Y. Wang and H. C. Chin, “An Empirically-Calibrated Ship Domain as a Safety Criterion for Navigation in Confined Waters,” *J. Navig.*, vol. 69, no. 2, pp. 257–276, 2016.

[51] M. G. Hansen, T. K. Jensen, T. Lehnschioler, K. Melchild, F. M. Rasmussen, and F. Ennemark, “Empirical Ship Domain based on AIS Data,” *J. Navig.*, vol. 66, no. 06, pp. 931–940, 2013.

[52] R. Szlapczynski and J. Szlapczynska, “Review of ship safety domains: Models and applications,” *Ocean Eng.*, vol. 145, pp. 277–289, 2017.

[53] R. Szłapczyński and T. Niksa-Rynkiewicz, “A framework of a ship domain-based near-miss detection method using mamdani neuro-fuzzy classification,” *Polish Marit. Res.*, vol. 25, no. s1, pp. 14–21, 2018.

[54] X. Wu, A. L. Mehta, V. A. Zaloom, and B. N. Craig, “Analysis of waterway transportation in Southeast Texas waterway based on AIS data,” *Ocean Eng.*, vol. 121, pp. 196–209, 2016.

[55] H.-J. Lee and K. P. Rhee, “Development of collision avoidance system by using expert system and search algorithm,” *Int. Shipbuild. Prog.*, vol. 48, no. 3, pp. 197–212, 2001.

[56] Y. Zhao, W. Li, and P. Shi, “A real-time collision avoidance learning system for Unmanned Surface Vessels,” *Neurocomputing*, vol. 182, pp. 255–266, 2016.

[57] L. Gang, Y. Wang, Y. Sun, L. Zhou, and M. Zhang, “Estimation of vessel collision risk index based on support vector machine,” *Adv. Mech. Eng.*, vol. 8, no. 11, p. 1687814016671250, 2016.

[58] J.-H. Ahn, K.-P. Rhee, and Y.-J. You, “A study on the collision avoidance of a ship using neural networks and fuzzy logic,” *Appl. Ocean Res.*, vol. 37, pp. 162–173, 2012.

[59] B. Li and F.-W. Pang, “An approach of vessel collision risk assessment based on the D–S evidence theory,” *Ocean Eng.*, vol. 74, pp. 16–21, 2013.

[60] M. Gil, J. Montewka, P. Krata, and T. Hinz, “Ship stability-related effects on a critical distance of collision evasive action,” in *Proceedings of the 17th Internationa l Ship Stability Workshop. Presented at the 17th International Ship Stabilit y Workshop, Helsinki, Finland*, 2019, pp. 231–238.

[61] M. Gil, J. Montewka, P. Krata, T. Hinz, and S. Hirdaris, “Determination of the dynamic critical maneuvering area in an encounter between two vessels: Operation with negligible environmental disruption,” *Ocean Eng.*, vol. 213, p. 107709, 2020.

[62] S. Li, J. Zhou, and Y. Zhang, “Research of vessel traffic safety in ship routeing precautionary areas based on navigational traffic conflict technique,” *J. Navig.*, vol. 68, no. 3, pp. 589–601, 2015.

[63] J. Montewka, T. Hinz, P. Kujala, and J. Matusiak, “Probability modelling of vessel collisions,” *Reliab. Eng. Syst. Saf.*, vol. 95, no. 5, pp. 573–589, 2010.

[64] F. Goerlandt, J. Montewka, V. Kuzmin, and P. Kujala, “A risk-informed ship collision alert system: framework and application,” *Saf. Sci.*, vol. 77, pp. 182–204, 2015.

[65] Z. Fang, H. Yu, R. Ke, S.-L. Shaw, and G. Peng, “Automatic identification system-based approach for assessing the near-miss collision risk dynamics of ships in ports,” *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 2, pp. 534–543, 2018.

[66] W. Zhang, X. Feng, Y. Qi, S. Feng, Y. Zhang, and Y. Wang, “Towards a model of regional vessel near-miss collision risk assessment for open waters based on AIS data,” *J. Navig.*, vol. 72, no. 6, pp. 1449–1468, 2019.

[67] A. S. Lenart, “Analysis of collision threat parameters and criteria,” *J. Navig.*, vol. 68, no. 5, pp. 887–896, 2015.

[68] J. Larson, M. Bruch, and J. Ebken, “Autonomous navigation and obstacle avoidance for unmanned surface vehicles,” in *Unmanned systems technology VIII*, 2006, vol. 6230, p. 623007.

[69] Y. Huang, P. van Gelder, and Y. Wen, “Velocity obstacle algorithms for collision prevention at sea,” *Ocean Eng.*, vol. 151, pp. 308–321, 2018.

[70] Y. Huang and P. van Gelder, “Time-Varying Risk Measurement for Ship Collision Prevention,” *Risk Anal.*, vol. 40, no. 1, pp. 24–42, 2020.

[71] Y. Huang, L. Chen, and P. van Gelder, “Generalized velocity obstacle algorithm for preventing ship collisions at sea,” *Ocean Eng.*, vol. 173, pp. 142–156, 2019.

[72] Y. Wen, Y. Huang, C. Zhou, J. Yang, C. Xiao, and X. Wu, “Modelling of marine traffic flow complexity,” *Ocean Eng.*, vol. 104, pp. 500–510, 2015.

[73] K. Baek and H. Bang, “ADS-B based trajectory prediction and conflict detection for air traffic management,” *Int. J. Aeronaut. Sp. Sci.*, vol. 13, no. 3, pp. 377–385, 2012.

[74] I. Hwang and C. E. Seah, “Intent-based probabilistic conflict detection for the next generation air transportation system,” *Proc. IEEE*, vol. 96, no. 12, pp. 2040–2059, 2008.

[75] W. Liu and I. Hwang, “Probabilistic trajectory prediction and conflict detection for air traffic control,” *J. Guid. Control. Dyn.*, vol. 34, no. 6, pp. 1779–1789, 2011.

[76] R. A. Paielli and H. Erzberger, “Conflict probability estimation generalized to non-level flight,” *Air Traffic Control Q.*, vol. 7, no. 3, pp. 195–222, 1999.

[77] V. P. Jilkov, J. H. Ledet, and X. R. Li, “Multiple model method for aircraft conflict detection and resolution in intent and weather uncertainty,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 55, no. 2, pp. 1004–1020, 2018.

[78] L. Yang, J. H. Yang, J. Kuchar, and E. Feron, “A real-time Monte Carlo implementation for computing probability of conflict,” in *AIAA Guidance, Navigation, and Control Conference and Exhibit*, 2004, p. 4876.

[79] S. Hao, Y. Zhang, S. Cheng, R. Liu, and Z. Xing, “Probabilistic multi-aircraft conflict detection approach for trajectory-based operation,” *Transp. Res. Part C Emerg. Technol.*, vol. 95, pp. 698–712, 2018.

[80] M. Prandini and J. Hu, “A stochastic approximation method for reachability computations,” in *Stochastic Hybrid Systems*, Springer, 2006, pp. 107–139.

[81] H. A. P. Blom, G. J. Bakker, P. J. G. Blanker, J. Daams, M. H. C. Everdij, and M. B. Klompstra, “Accident risk assessment for advanced air traffic management,” 2001.

[82] M. Kochenderfer, D. Griffith, and J. Olszta, “On estimating mid-air collision risk,” in *10th AIAA aviation technology, integration, and operations (ATIO) conference*, 2010, p. 9333.

[83] M. Prandini and O. J. Watkins, “Probabilistic aircraft conflict detection,” *HYBRIDGE, IST-2001*, vol. 32460, no. 2, 2005.

[84] M. Mitici and H. A. P. Blom, “Mathematical models for air traffic conflict and collision probability estimation,” *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 3, pp. 1052–1068, 2018.

[85] Y. Cho and J. Kim, “Collision probability assessment between surface ships considering maneuver intentions,” in *OCEANS 2017-Aberdeen*, 2017, pp. 1–5.

[86] J. L. Yepes, I. Hwang, and M. Rotea, “New algorithms for aircraft intent inference and trajectory prediction,” *J. Guid. Control. Dyn.*, vol. 30, no. 2, pp. 370–382, 2007.

[87] R. A. Paielli, H. Erzberger, D. Chiu, and K. R. Heere, “Tactical conflict alerting aid for air traffic controllers,” *J. Guid. Control. Dyn.*, vol. 32, no. 1, pp. 184–193, 2009.

[88] Y. Yang, J. Zhang, K.-Q. Cai, and M. Prandini, “Multi-aircraft conflict detection and resolution based on probabilistic reach sets,” *IEEE Trans. Control Syst. Technol.*, vol. 25, no. 1, pp. 309–316, 2016.

[89] H. Rong, A. P. Teixeira, and C. G. Soares, “Data mining approach to shipping route characterization and anomaly detection based on AIS data,” *Ocean Eng.*, vol. 198, p. 106936, 2020.

[90] Q. Yu, K. Liu, A. P. Teixeira, and C. G. Soares, “Assessment of the Influence of Offshore Wind Farms on Ship Traffic Flow Based on AIS Data,” *J. Navig.*, vol. 73, no. 1, pp. 131–148, 2020.

[91] X. Xin, K. Liu, X. Yang, Z. Yuan, and J. Zhang, “A simulation model for ship navigation in the ‘Xiazhimen’ waterway based on statistical analysis of AIS data,” *Ocean Eng.*, vol. 180, pp. 279–289, 2019.

[92] H. Li, J. Liu, Z. Yang, R. W. Liu, K. Wu, and Y. Wan, “Adaptively constrained dynamic time warping for time series classification and clustering,” *Inf. Sci. (Ny).*, 2020.

[93] H. Li, J. Liu, K. Wu, Z. Yang, R. W. Liu, and N. Xiong, “Spatio-temporal vessel trajectory clustering based on data mapping and density,” *IEEE Access*, vol. 6, pp. 58939–58954, 2018.

[94] M. Zhang, J. Montewka, T. Manderbacka, P. Kujala, and S. Hirdaris, “Analysis of the Grounding Avoidance Behavior of a Ro-Pax Ship in the Gulf of Finland using Big Data,” in *The 30th International Ocean and Polar Engineering Conference*, 2020.

[95] C. Liu and X. Chen, “Inference of single vessel behaviour with incomplete satellite-based AIS data,” *J. Navig.*,

vol. 66, no. 6, p. 813, 2013.

[96] L. Chen, Y. Huang, H. Zheng, H. Hopman, and R. Negenborn, “Cooperative multi-vessel systems in urban waterway networks,” *IEEE Trans. Intell. Transp. Syst.*, 2019.

[97] L. Chen, H. Hopman, and R. R. Negenborn, “Distributed model predictive control for vessel train formations of cooperative multi-vessel systems,” *Transp. Res. Part C Emerg. Technol.*, vol. 92, pp. 101–118, 2018.

[98] D. H. Douglas and T. K. Peucker, “Algorithms for the reduction of the number of points required to represent a digitized line or its caricature,” *Cartogr. Int. J. Geogr. Inf. geovisualization*, vol. 10, no. 2, pp. 112–122, 1973.

[99] S. Zhang, Z. Liu, Y. Cai, Z. Wu, and G. Shi, “AIS trajectories simplification and threshold determination,” *J. Navig.*, vol. 69, no. 4, pp. 729–744, 2016.

[100] L. Zhao and G. Shi, “A method for simplifying ship trajectory based on improved Douglas–Peucker algorithm,” *Ocean Eng.*, vol. 166, pp. 37–46, 2018.

[101] L. Du, F. Goerlandt, O. A. V. Banda, Y. Huang, Y. Wen, and P. Kujala, “Improving stand-on ship’s situational awareness by estimating the intention of the give-way ship,” *Ocean Eng.*, vol. 201, p. 107110, 2020.

[102] M. Chen, K. Liu, J. Ma, Y. Gu, Z. Dong, and C. Liu, “SWIM: Speed-aware WiFi-based Passive Indoor Localization for Mobile Ship Environment,” *IEEE Trans. Mob. Comput.*, 2019.

[103] B. W. Silverman, *Density estimation for statistics and data analysis*, vol. 26. CRC press, 1986.

[104] L. Kang, Q. Meng, and Q. Liu, “Fundamental diagram of ship traffic in the Singapore Strait,” *Ocean Eng.*, vol. 147, pp. 340–354, 2018.

[105] L. Zhao, G. Shi, and J. Yang, “Ship trajectories pre-processing based on AIS data,” *J. Navig.*, vol. 71, no. 5, pp. 1210–1230, 2018.

[106] L. Zhang, Q. Meng, and T. F. Fwa, “Big AIS data based spatial-temporal analyses of ship traffic in Singapore port waters,” *Transp. Res. Part E Logist. Transp. Rev.*, vol. 129, pp. 287–304, 2019.

[107] M. Hollander, D. A. Wolfe, and E. Chicken, *Nonparametric statistical methods*, vol. 751. John Wiley & Sons, 2013.

[108] C. M. Jarque and A. K. Bera, “Efficient tests for normality, homoscedasticity and serial independence of regression residuals,” *Econ. Lett.*, vol. 6, no. 3, pp. 255–259, 1980.

[109] F. J. Massey Jr, “The Kolmogorov-Smirnov test for goodness of fit,” *J. Am. Stat. Assoc.*, vol. 46, no. 253, pp. 68–78, 1951.

[110] B. Ożoga and J. Montewka, “Towards a decision support system for maritime navigation on heavily trafficked basins,” *Ocean Eng.*, vol. 159, pp. 88–97, 2018.

[111] X. Wang, Z. Liu, and Y. Cai, “The ship maneuverability based collision avoidance dynamic support system in close-quarters situation,” *Ocean Eng.*, vol. 146, pp. 486–497, 2017.

[112] R. Zhen, M. Riveiro, and Y. Jin, “A novel analytic framework of real-time multi-vessel collision risk assessment for maritime traffic surveillance,” *Ocean Eng.*, vol. 145, pp. 492–501, 2017.